

# Towards an Open Source Analysis Toolbox for Street Network Comparison: Indicators, Tools and Results of a Comparison of OSM and the Official Austrian Reference Graph

Anita Graser, Markus Straub, Melitta Dragaschnig

## Abstract

This paper presents a novel open source toolbox for street network comparison based on the Sextante geoprocessing framework for the open source Geographic Information System Quantum GIS (QGIS). In the spirit of open science, the toolbox enables researchers worldwide to assess the quality of street networks such as OpenStreetMap (OSM) by calculating key performance indicators commonly used in street network comparison studies. Additionally, we suggest two new performance indicators for turn restriction and one-way street comparisons specifically aimed at testing street network quality for routing. We demonstrate the use of this toolbox by comparing OSM and the official Austrian reference graph “Graph Integration Platform” (GIP) in the greater Vienna region.

## Keywords

OpenStreetMap, Volunteered Geographic Information (VGI), quality assessment, open science, open source, Austria

## 1 Introduction

Concerns about the quality and reliability of Volunteered Geographic Information (VGI), including those of OSM, have lead to the publication of numerous papers following the first systematic analysis of the quality of OSM in the UK (Haklay 2010). Published research on OSM quality includes work on OSM-internal validity and topology checks, such as in Neis et al. (2012), as well as a variety of papers comparing OSM to other free and commercial datasets worldwide. The following paragraphs present an overview of related work.

Haklay (2010) presents an approach for determining the positional accuracy of OSM compared to the Ordnance Survey Meridian 2 street network dataset based on work by Goodchild & Hunter (1997). His approach “*relies on a comparison with a representation of higher accuracy, and estimates the percentage of the total*

*length of the low accuracy representation that is within a specified distance of the high accuracy representation.*” Comparisons were carried out for UK motorways and found an average overlap of nearly 80 %.

Haklay (2010) further presents a test for completeness by comparing network length by analysis cell. His analysis was adopted in many subsequent papers for the UK (Ather 2009), Ireland (Ciepluch et al. 2011), the US (Zielstra & Hochmair 2011, 2012), Germany (Zielstra & Zipf 2010, Neis et al. 2012) and Austria (Thaller 2009), comparing OSM to Ordnance Survey Ireland, TeleAtlas/TomTom MultiNet, Nokias NAVTEQ and TIGER data. Zielstra & Zipf (2010) and Zielstra & Hochmair (2011) compared network lengths in different cities, for both the full street network as well as the pedestrian-related network only.

Ather (2009) extends Haklay’s work and additionally compares completeness of street names based on the percentage of network length with and without names in selected areas of London – an analysis that can also be found in Neis et al. (2012) for Germany. Thaller (2009) compares the number of distinct street names instead. Neis et al. (2012) additionally compare the number of turn restrictions found in OSM and TomTom by street category and city type. Results show that the number of turn restrictions in OSM is significantly lower than in TomTom MultiNet (21,000 compared to 176,000 in Germany, June 2011).

Ludwig et al. (2011) and Koukoletsos et al. (2012) describe methods based on a match between the street objects of OSM and objects in a reference dataset – Navteq and Ordnance Survey, respectively – which enables them to compare street object geometries and attributes.

The analyses presented in the above OSM comparison papers have been performed using a variety of tools – often involving custom code. Haklay (2010) uses MapInfo for clipping and Manifold GIS for analysis. Ather (2009) uses ArcGIS, Manifold GIS and MapInfo. Ciepluch et al. (2011) use QGIS, PostGIS and custom PHP code to draw comparisons. Zielstra & Hochmair (2011) use ArcGIS 9.3.1 and ArcGIS 10 in Zielstra & Hochmair (2012). Neis et al. (2012) use PostGIS, OSMOSIS and custom Java tools developed using GeoTools.

The British Royal Society recommends that “*scientists should communicate data they collect and models they create via methods that allow free and open access and in ways that are intelligible, assessable and usable for other specialists in the same or linked fields wherever they are in the world*” (Royal Society 2012). In this spirit, this paper presents algorithms and models built on the QGIS Sextante geoprocessing framework for calculating the following performance indicators which are found in related work: positional accuracy, network length and attribute completeness. These tools can be used without further software development knowledge from within QGIS, and are designed for researchers worldwide to freely use and improve. Additionally, we propose two new test methods specifically aimed at testing street network quality for routing by use of dedicated turn restriction and one-way street comparisons. We further demonstrate the use of our toolbox by comparing OSM and the official Austrian reference graph “Graph Integration Platform” (GIP) in the greater Vienna region. Finally, we evaluate the

comparison results by performing on-site checks with respect to real-world traffic situations for selected areas.

Section 2 presents algorithms and Sextante geoprocessing models using in this work. Section 3 compares OSM and the graph export for motorized private traffic of the official Austrian reference street graph GIP. The results of this comparison are further compared to the real-world situation in selected regions. Section 4 summarizes the results of this study and offers an outlook on future work.

## 2 Analysis tools

We developed analysis tools for street network comparison as geoprocessing models using QGIS 1.8 (Quantum GIS Development Team 2013) with Sextante 1.0.8 geoprocessing framework (Sextante 2013). The analysis tools use existing functionality as well as new scripts created specifically for this study. All code is open source (under the GNU General Public License) and are available on Github together with instructions of how to install the tools (<https://github.com/anitagraser/QGIS-Sextante/tree/master/1.0.8/street-network-comparison>). All models include reprojection steps to ensure a common analysis coordinate reference system (CRS). These CRS settings have to be adapted by the user before the tools can be applied to different geographic regions.

The first three geoprocessing models calculate positional accuracy (Section 2.1), network length comparison (Section 2.2), and attribute completeness (Section 2.3). These geoprocessing tools can be used with standard OSM exports such as the Shapefile exports provided by Geofabrik (2013). In Sections 2.4 and 2.5, we present two new test methods for turn restrictions and one-way streets which are specifically aimed at testing street network quality for routing. These geoprocessing models test for matches in turn restrictions and one-way streets in the reference graph by use of routing and therefore require a routable graph. Preparation of routable graphs based on raw OSM data is discussed in Section 3.1. Graph preparation and routing using Dijkstra's shortest path algorithm (Dijkstra 1959) are performed outside QGIS. Routing results are exchanged using comma-separated text files with geometries in Well-known Text (WKT) format but any other file format readable by QGIS can be used as well.

### 2.1 Positional accuracy

Based on Goodchild & Hunter (1997), the geoprocessing model as depicted in Figure 1 calculates the percentage of the total length of one graph that is within a specified distance of the reference graph in user-defined regions. The required inputs are: the graph under evaluation, a reference graph and a layer defining analysis regions (regular cells or irregular regions), information about the region ID field and the selected distance (buffer size) in meters. The reference graph is buffered and the buffers dissolved (using "Difference" tool). The length of the graph within

the buffer is compared to the length in the whole region (calculated by “Sum line lengths” tool). The tools “join on attribute” and “division” were created specifically for this study since Sextante 1.0.8 does not feature a join tool and the default Sextante field calculator cannot handle divisions by zero, which do occur in regions that do not contain any graph edges. (“Division” outputs NULL in case of division by zero.) The model outputs a new region layer with additional attributes for total graph length, length of the graph within the buffered reference graph and the ratio of these two values.

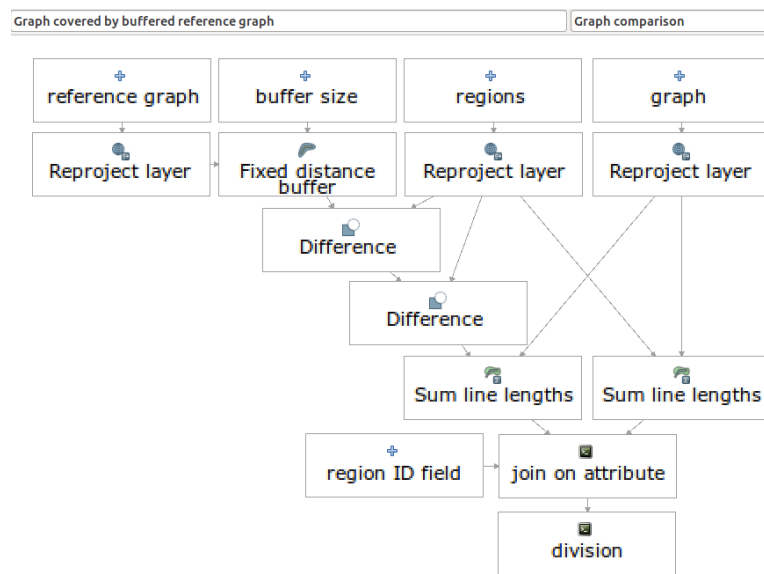


Figure 1: Positional accuracy test model

## 2.2 Network length comparison

Network length comparison is the most common test for completeness of street networks used in Haklay (2010) and numerous subsequent publications. Figure 2 shows the model for simple comparison of total graph length in user-defined regions. The required inputs are: two graphs as well as a region layer and information about the region ID field. The network length is calculated using “Sum line lengths” tool. The model outputs a new region layer with appended attributes for both graphs’ length per cell as well as the difference of these two measures.

We developed a second model to calculate the length of subsets (subnetworks) of the input graph (Figure 3). The output of this model enables analyses of different street classes, i.e. can be used to determine the length of the pedestrian-only network compared to the full network. This model is also used as a building block in the following model which evaluates attribute completeness. For each distinct value in the user-defined classification field, the length of the corresponding subnetwork is calculated and added to the output layers attribute table. To achieve

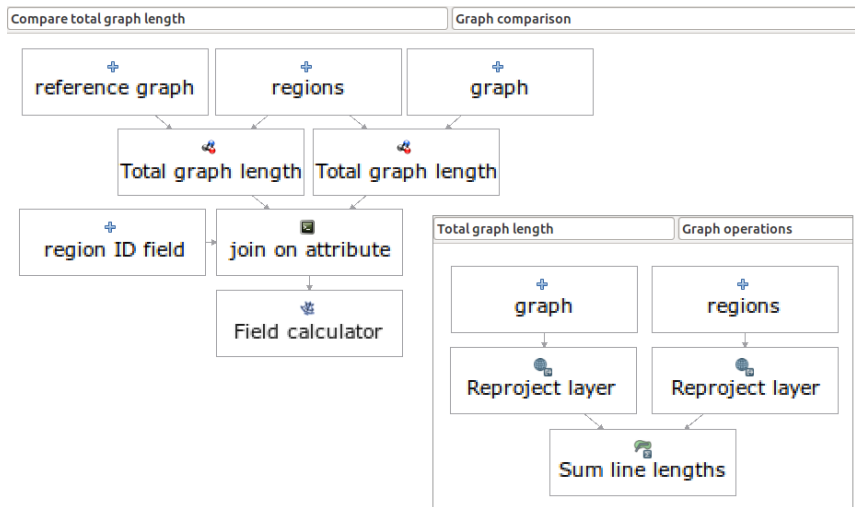


Figure 2: Network length comparison model. Note that the main model “Compare total graph length” contains two instances of the submodel “Total graph length” which was created separately to improve the modularity of the toolbox.

this, “Sum line lengths” was extended to include additional steps to handle multiple classes and named “Sum line lengths by class”.

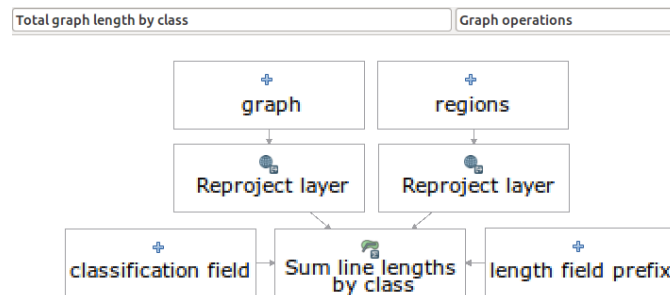


Figure 3: Subnetwork length calculation model

### 2.3 Attribute completeness

Completeness of attributes such as street name (Ather 2009) and speed is another important performance indicator due to the impact these attributes have on a wide range of applications, from visualizing the street network to routing and navigation. Figure 4 shows the model used to determine attribute completeness. It can be interpreted as an extended version of the “Total graph length by class” model: One class consists of all edges containing an attribute value in the classification field, the other class comprises edges without such a value. The classification is performed by the “classify null notnull” tool we developed. The model takes a

graph, a classification field and a region layer as input and outputs a new region layer with appended attributes for the length of the two subgraphs.

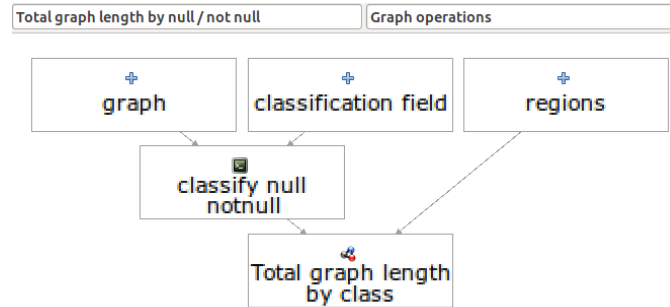


Figure 4: Model calculating length of subgraphs with and without attribute of interest

## 2.4 Turn restriction comparison

Due to modelling differences between street networks – which are discussed in more detail in the case study for OSM and GIP in Section 3 – quality assessments based on turn restriction counts may be unreliable. We therefore propose a novel routing-based approach: Starting from one graph, a list of all forbidden turn maneuvers is compiled including a spatial representation for each forbidden turn maneuver: a line geometry starting from the center of the source edge of the turn restriction and ending at the center of the target edge as shown by the black arrows in Figure 5.

To test whether a certain turn restriction is also modelled in the comparison graph, a shortest-path route is calculated on the comparison graph, starting at the first point of the forbidden turn and ending at the last point of the forbidden turn. If the turn restriction is modelled correctly in the comparison graph, the resulting route geometry will differ from the forbidden turn geometry (Figure 5: left), because the router has to find a way around the turn restriction. In contrast, a missing turn restriction will result in a route which is very similar to the forbidden turn (Figure 5: right).

Similarity between forbidden turn and resulting route is determined using Hausdorff distance (Hausdorff 1914). Intuitively, Hausdorff distance measures the largest of all minimal distances between any pair of points within the two line geometries to be compared. Hausdorff distance was implemented as an alternative to the buffer method for positional accuracy as presented by Goodchild & Hunter (1997), since it can be computed once to obtain Hausdorff distance values which can be further categorized into similar and dissimilar pairs of forbidden turn and route instead of having to rerun the buffer operations for each new threshold value.

For a certain street, the associated edge geometries in different graphs may differ slightly. For the analysis presented in the following case study, we therefore



Figure 5: left: correctly modelled turn restriction; right: missing restriction on the right. (narrow black arrows: forbidden turn maneuvers; colored arrows: routes generated by routing on the comparison graph)

used a Hausdorff distance threshold of ten meters to account for such differences. Pairs of forbidden turn geometries and resulting routes with Hausdorff distances below the specified threshold are regarded as similar. In other words, the routing result is very similar to the prohibited turn maneuver geometry, and we therefore consider the turn restriction missing. Hausdorff distances above the threshold indicate a correctly modelled turn restriction.

Figure 6 shows the model used to calculate Hausdorff distances between forbidden turns in one layer and routes generated on the comparison graph in another layer. Besides these two input layers, it is necessary to specify join fields which are used to determine the two geometries the Hausdorff distance should be calculated for. Since the Hausdorff distance tool implemented for this paper calculates distances between line vertices, we implemented an extra step to densify the input geometries and add extra vertices at regular intervals. The model outputs a new layer containing geometries and attributes of the first input route layer and an additional attribute containing the Hausdorff distance between this geometry and the matching geometry in the second layer.

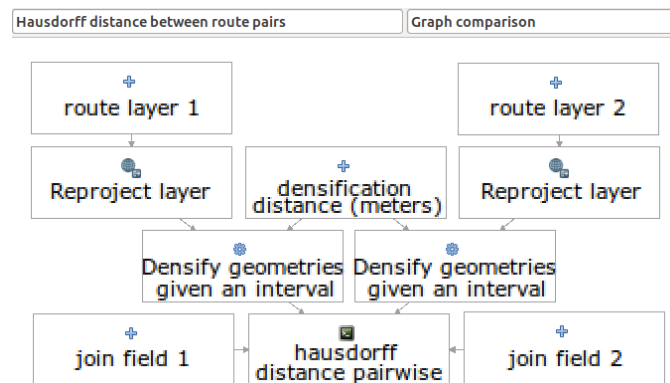


Figure 6: Model calculating Hausdorff distance between pairs of geometries

## 2.5 One-way street comparison

Due to modelling differences between street networks – which are discussed in more detail in the case study in Section 3 – quality assessments solely based on one-way street counts may be unreliable. We therefore propose a new routing-based approach: Starting from one graph, a list of all edges marked as one-way streets is compiled. This list includes a spatial representation of the one-way edge consisting of a short section of ten meters length extracted from the center of the one-way edge.

Between the end points of this extracted section a shortest-path route is calculated on the comparison graph against the permitted driving (one-way) direction. If the one-way information in both graphs matches, the router has to find a way around the driving restriction. This will result in a route which is much longer than the short ten meter extract (Figure 7: left). Conversely, if one-way information is missing the shortest possible route which will be very close to ten meters (Figure 7: right). As in the turn restriction comparison, the one-way analysis presented in the following case study uses a route length difference of ten meters which results in a threshold of twenty meters since all forbidden geometries are ten meters long. Only one-way streets with an associated route longer than this threshold are considered to be correct. If the route is shorter than twenty meters the one-way information is considered to be wrong or missing.



Figure 7: left: correctly modelled one-way street; right: missing one-way restriction. (narrow black arrows: extracted section against the one-way direction; wide colored arrows: routes generated on the comparison graph)

Figure 8 shows the model used to calculate and compare route lengths. It can be used to compare one-way sections with the corresponding routes but is not limited to this use case. The model takes two route layers (e.g. a layer of one-way streets and a layer of corresponding routes for our purposes) as input and outputs a new layer containing geometries of the first input route layer as well as attributes and length measures of both input layers and the length difference.



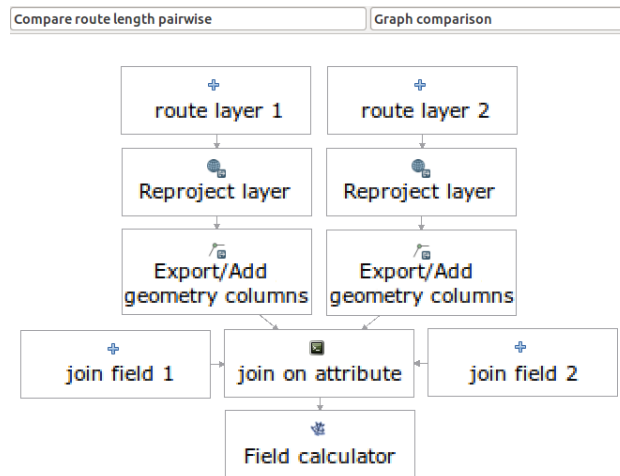


Figure 8: Model calculating route length and length difference

### 3 Case study: Comparing OSM and GIP

We applied the geoprocessing models presented in Section 2 to compare OSM and the graph export for motorized private traffic of the official Austrian reference street graph GIP within the greater Vienna region. The datasets used for this comparison are the OSM export provided by Geofabrik.de for December, 12th 2012 and the GIP export from December, 18th 2012 for routing motorized traffic (called “MIV export”). Section 3.1 describes the different approaches to street network modelling used in OSM and GIP followed by the steps necessary to generate a routing graph in Section 3.2, which are important to interpret the results presented in Section 3.3. In Section 3.4 We further compared the results of this comparison to the real-world situation in selected regions.

#### 3.1 Street network modelling in OSM and GIP

Table 1 presents a short overview of the different street network modelling approaches. While GIP tends to explicitly define features such as driving **permissions** and permitted turn relations, OSM tends to use conventions and explicit **restrictions**.

The GIP **street network** is modelled using nodes and edges which are connected through explicit turn relations. Without a turn relation, edges sharing a common intersection node are not considered to be connected. The OSM street network is also represented using nodes and edges but, unlike GIP edges, OSM edges are not necessarily split at each intersection. Instead, edges are considered to be connected if they share a common node at the point of their intersection. Therefore OSM is not routable without appropriate preprocessing.

|                              | <b>GIP</b>                                               | <b>OSM</b>                                                                  |
|------------------------------|----------------------------------------------------------|-----------------------------------------------------------------------------|
| <b>exchange format</b>       | IDF (CSV)                                                | OSM (XML), PBF (binary)                                                     |
| <b>graph structure</b>       | nodes and edges connected by driving permissions         | nodes and edges connected by shared nodes                                   |
| <b>street classification</b> | extended FRC definition based on TeleAtlas FRC 0 to 8    | 10 classes for motorized traffic                                            |
| <b>driving permissions</b>   | for each mode of transport and driving direction         | convention and explicit restrictions, generally for both driving directions |
| <b>one-way streets</b>       | derived from driving permissions                         | dedicated attribute                                                         |
| <b>speed limits</b>          | average speed limited by legal speed limit               | legal speed limit                                                           |
| <b>turn restrictions</b>     | derived from missing turn relations and u-turn attribute | explicit turn restriction relations                                         |

Table 1: Approaches to street network modelling OSM and GIP

In GIP, **access permissions** are explicitly specified as either allowed or forbidden for each defined mode of transport. In OSM, country-specific default permissions are applied if nothing else is specified. A number of tags are used to define non-default driving permissions. Permissions for modes of transport are evaluated using a hierarchy; i.e. the tag combination `vehicle=no` and `bicycle=yes` evaluates to a ban on all vehicles except bicycles.

**GIP turn restrictions** are modelled implicitly through missing turn relations and can be extracted from the routing graph by testing for turn relations at each intersection. A turn restriction is registered if no corresponding turn relation is found. In OSM, all turn manoeuvres are allowed at an intersection if there is no explicit restriction relation specified. A restriction relation always consists of two edges (from and to) and a connecting object which in most cases is a node, but may also be an edge. Relation restrictions can be either bans such as “no left turn allowed”, or constraints such as “only straight on”.

GIP contains 2,500 turn restrictions while OSM contains only 691 in the case study area depicted in Figure 9. These values cannot be compared directly because the same situation can be modelled using fewer relations in OSM than in GIP. A situation where street users are only allowed to drive straight on at an intersection of four streets can be modelled using a single “only straight on” constraint in OSM, while two turn restrictions (for turning left and turning right) are required in the GIP graph. In the following comparison, we therefore systematically calculate routes for each GIP turn restriction to test if it is also modelled in OSM. A similar approach is used to compare GIP and OSM one-way streets.

Streets are **classified by importance** in both GIP (functional road class FRC)

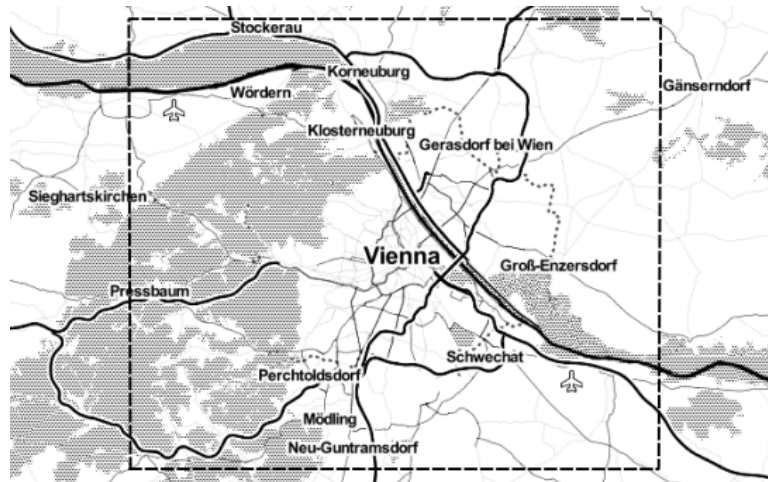


Figure 9: Overview of the study area (map by Stamen, data by OSM)

and OSM (highway tag). The FRC definition used in GIP is based on the definition by TeleAtlas (with FRC from 0 to 8) and has been extended to include more classes (e.g. 11: farm road). OSM defines ten street classes for motorized traffic. Additional classes for connecting street types such as “motorway\_link” to indicate a slip street, exist for the top five street classes. Table 2 shows the matching between OSM highway classes and GIP FRC which was used for the following comparisons. This matching is the result of a manual comparison of the streets contained in each class. Five groups (A to E) were identified: A (motorway and FRC0), B (trunk/primary and FRC1/2), C (secondary/tertiary and FRC3/4), D (unclassified/living\_street/service/residential and FRC5/6/7/8) and E (track and FRC11/108).

|                      | 0   | 1 | 2 | 3   | 4 | 5       | 6 | 7 | 8 | 11 or 108 |
|----------------------|-----|---|---|-----|---|---------|---|---|---|-----------|
| <b>motorway</b>      | A   |   |   |     |   |         |   |   |   |           |
| <b>trunk</b>         | B B |   |   |     |   |         |   |   |   |           |
| <b>primary</b>       | B B |   |   |     |   |         |   |   |   |           |
| <b>secondary</b>     |     |   |   | C C |   |         |   |   |   |           |
| <b>tertiary</b>      |     |   |   | C C |   |         |   |   |   |           |
| <b>unclassified</b>  |     |   |   |     |   | D D D   |   |   |   |           |
| <b>living_street</b> |     |   |   |     |   | D D D   |   |   |   |           |
| <b>service</b>       |     |   |   |     |   | D D D   |   |   |   |           |
| <b>residential</b>   |     |   |   |     |   | D D D D |   |   |   |           |
| <b>track</b>         |     |   |   |     |   |         |   |   |   | E         |

Table 2: Matching between OSM highway and GIP FRC

In both street graphs, edges can be assigned a speed value. In GIP, the specified speed value is the average speed, limited by the legal speed limit. In OSM, the

legal speed limit is specified and if no legal speed limit is defined, country-specific default limits can be applied.

## **3.2 Generating the routing graph**

For the one-way street and turn restriction comparison, it is necessary to route on the street graphs. Therefore, OSM raw data must be preprocessed following the data description outlined in Section 3.1: Edges that share nodes at their intersection have to be split up at the intersection node. Streets are grouped using OSM tags that indicate modes of transport (e.g. foot, bicycle, car). In order to create a correct multimodal routing graph, each driving direction of each street edge is labelled with a set of permitted modes of transport. Turn restrictions are created from all OSM tag combinations that define turn maneuvers. Every turn that is not explicitly forbidden is permitted. Turn restrictions are also attributed with the modes of transport they concern. These graph preparation steps are performed outside QGIS using a proprietary tool which is not part of the presented open source toolbox.

## **3.3 Results**

In the following comparisons, the GIP graph export is used as the reference graph which OSM is compared with. Consequently, features missing from the reference graph cannot be detected by this comparison. They can be discovered by other means, e.g. an on-site check (Section 3.4) or by exchanging reference graph and comparison graph.

### **3.3.1 Positional accuracy**

Based on the method for determining positional accuracy of linear features described in Goodchild & Hunter (1997), we compare OSM to GIP using buffers of different size around GIP motorways. Table 3 lists the percentage of the OSM motorway geometry within 10, 15 and 20 meters of the corresponding GIP motorway geometry. The results show that geometries in OSM and GIP match very well especially along the motorways' main corridors. Bigger differences can be found at motorway ramps. The worst values are registered for A5 Nordautobahn – a motorway which extends only a short distance into the analysis area. Therefore, ramp geometry differences influence the overall results more than in case of longer motorways.

### **3.3.2 Length comparison results**

A first comparison of total network length of OSM and GIP shows big differences (OSM: 11,039,626 m; GIP: 6,294,378 m). These differences are mostly due to dirt streets (marked in OSM using the tag `highway=track`) not being included in the GIP export we received. Figure 10 shows the length comparison for the street class

| motorway                          | 10 meters | 15 meters | 20 meters |
|-----------------------------------|-----------|-----------|-----------|
| A1 Westautobahn                   | 96.32 %   | 99.91 %   | 99.94 %   |
| A2 Südbahn                        | 98.69 %   | 100.00 %  | 100.00 %  |
| A4 Ostautobahn                    | 99.66 %   | 99.69 %   | 99.71 %   |
| A5 Nordautobahn                   | 67.50 %   | 86.69 %   | 94.99 %   |
| A21 Wiener Außenring-Autobahn     | 97.24 %   | 100.00 %  | 100.00 %  |
| A22 Donauufer-Autobahn            | 96.64 %   | 98.12 %   | 98.41 %   |
| A23 Südosttangente                | 99.31 %   | 100.00 %  | 100.00 %  |
| S1 Wiener Außenring-Schnellstraße | 88.00 %   | 96.17 %   | 97.34 %   |

Table 3: Positional accuracy of motorways in the Vienna region

groups of Table 2, group E being empty in GIP because of the mentioned lack of dirt street information.

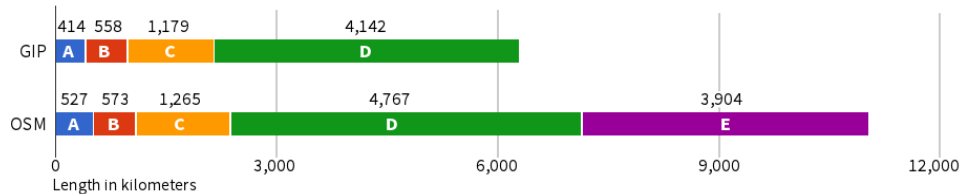


Figure 10: Comparison of subgraph length by street class group

The length difference between the groups GIP FRC0 and OSM motorway (group A) is due to motorway ramps being classified as part of the motorway group in case of OSM while they are classified FRC2 in GIP. Furthermore, a higher degree of generalization in GIP leads to a generally shorter network, since streets that are modelled as separate edges in OSM (due to e.g. traffic islands) can be represented by only one edge in the GIP export (see example in Figure 11).

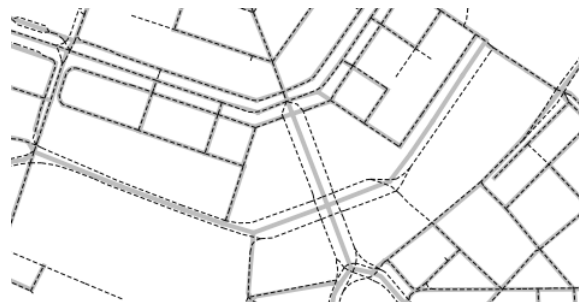


Figure 11: Details of OSM (narrow dashed black lines) and GIP (wide grey lines) at Schwarzenbergplatz

### 3.3.3 Attribute completeness results

Table 4 lists values of attribute completeness for street names (GIP name1 and OSM name) as well as speed (GIP speed\_tow\_car and OSM maxspeed). 94 % of GIP streets contain street name information. Especially within the urban area of Vienna, street names are close to complete as can be seen in Figure 12. In OSM, 78 % of streets include name information. This amounts to a difference of 400 km of named streets between the two street graphs.

Speed information is 100 % complete in GIP while only 43 % of the OSM network have a maxspeed value assigned. Especially lower-ranking streets in group D are affected by missing speed information: only 30 % of streets in this group contain a maxspeed value. The relatively low rate of streets with speed information might be explained by the convention in OSM that default speed limits for the given street class should be used when there is no explicit speed limit given.

|                                          | street name              |                          | speed                     |                          |
|------------------------------------------|--------------------------|--------------------------|---------------------------|--------------------------|
|                                          | GIP                      | OSM                      | GIP                       | OSM                      |
| <b>full network</b><br>(without group E) | 5,936,126 m<br>(94.31 %) | 5,539,584 m<br>(77.67 %) | 6,294,378 m<br>(100.00 %) | 3,078,633 m<br>(43.16 %) |
| <b>group A</b>                           | 402,039 m<br>(97.17 %)   | 359,336 m<br>(68.13 %)   | 413,768 m<br>(100.00 %)   | 394,877 m<br>(74.87 %)   |
| <b>group B</b>                           | 555,276 m<br>(99.60 %)   | 429,196 m<br>(74.87 %)   | 557,500 m<br>(100.00 %)   | 452,478 m<br>(78.94 %)   |
| <b>group C</b>                           | 1,152,662 m<br>(97.75 %) | 981,313 m<br>(77.59 %)   | 1,179,175 m<br>(100.00 %) | 813,189 m<br>(64.30 %)   |
| <b>group D</b>                           | 3,826,149 m<br>(92.33 %) | 3,769,739 m<br>(79.08 %) | 4,142,304 m<br>(100.00 %) | 1,418,089 m<br>(29.75 %) |

Table 4: Comparison of attribute completeness

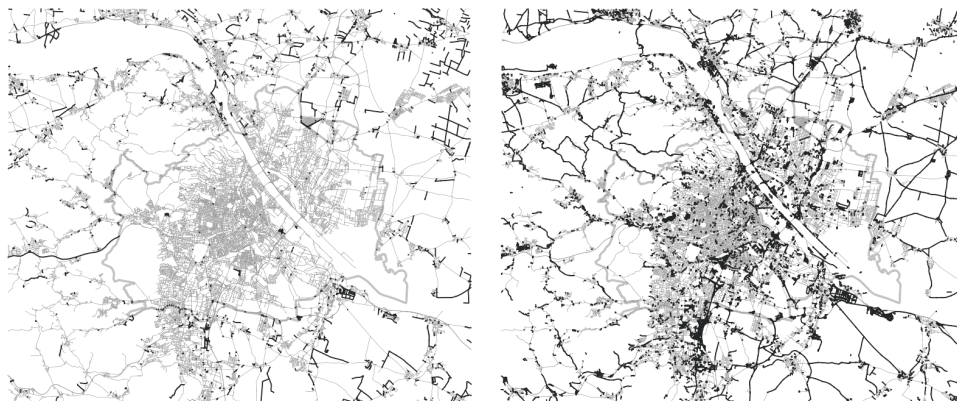


Figure 12: Streets without names (highlighted in black; GIP on the left, OSM on the right; Vienna city borders in bold grey)

### 3.3.4 Turn restriction comparison results

We systematically calculated routes for each GIP turn restriction to test if OSM contains a matching representation using the approach described in Section 2.4. Results show that 1,515 (60.6 %) of the 2,500 GIP turn restrictions in the Vienna region have a matching representation in OSM. Figure 13 shows the spatial distribution of turn restrictions in a regular 1 km<sup>2</sup> grid. Values displayed in the grid cells represent the absolute number of contained turn restrictions. Turn restrictions crossing the borders of multiple cells are counted multiple times, once for each containing cell. The ratio of matching turn restrictions is color-coded from white “no match” to black “perfect match”. 104 of the 381 cells contain at least one turn restriction match perfectly.

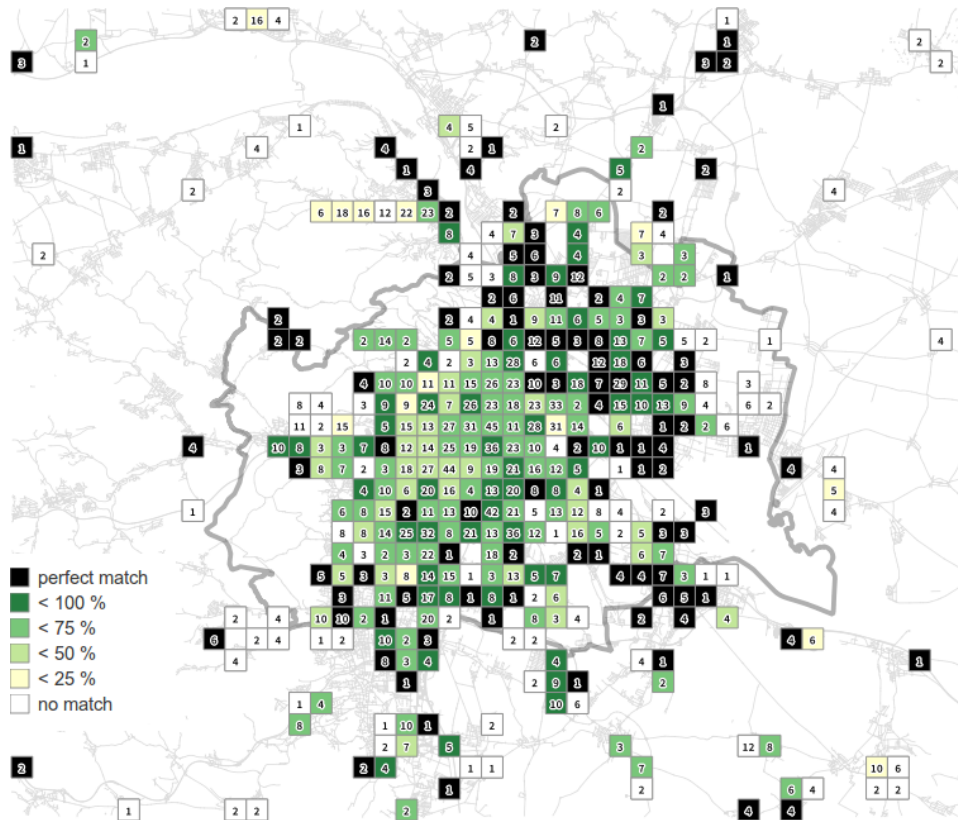


Figure 13: spatial distribution of turn restrictions (absolute values in cell) and ratio of matching turn restrictions (color/hue)

### 3.3.5 One-way street comparison results

We systematically calculated routes for each GIP one-way street to test if OSM contains a matching representation using the approach described in Section 2.5.

10,499 (87.8 %) of the 11,964 GIP one-way streets can be matched to a one-way street in OSM. Figure 14 shows the spatial distribution of one-way streets using the same representation as described for figure 13. The ratio of matching one-way streets tends to be higher in urban areas and lower in rural and industrial zones. 178 of the 581 cells that contain at least one one-way street match perfectly.

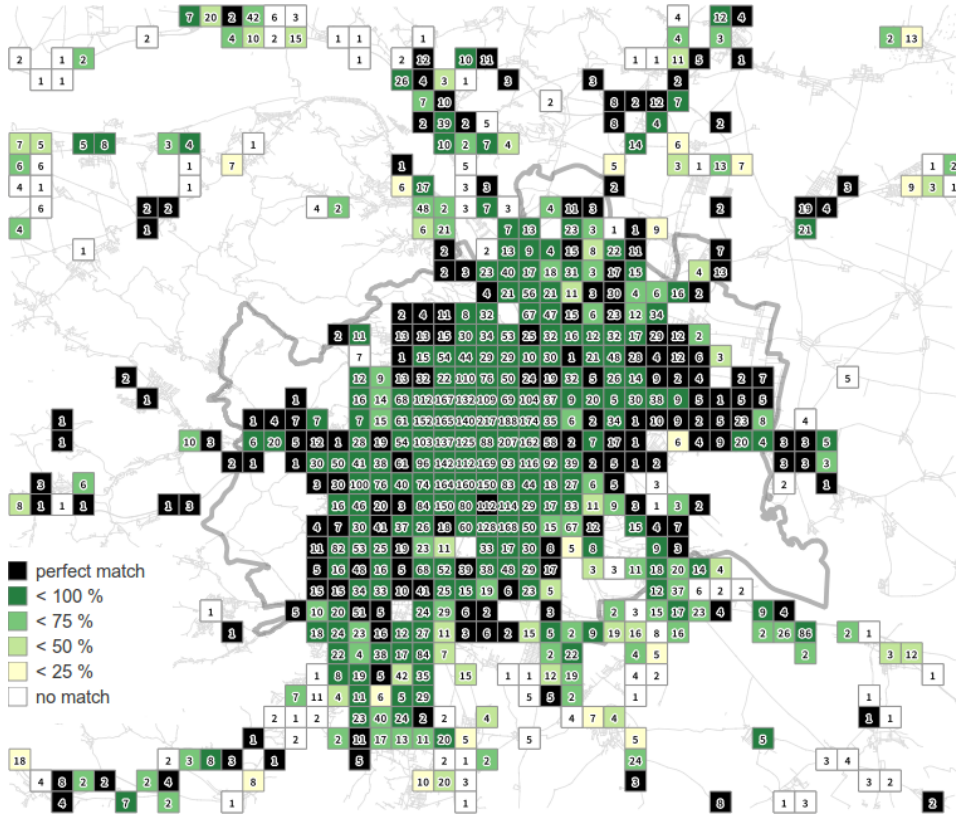


Figure 14: spatial distribution of one-way streets (absolute values in cell) and ratio of matching one-way streets (color)

### 3.4 On-site verification

For selected regions in the city of Vienna, we performed an on-site evaluation of the differences found by comparing OSM and GIP. The 9th district was chosen as a representative for the inner urban area, where OSM and GIP tend to match well. Additionally, the so-called Stuwerviertel was evaluated because it is a hotspot of differing turn restrictions. Parts of the 17th district between Sandleitengasse and Roggendorfgasse were evaluated on-site as well because of the high number of differing one-way streets located in this area.

In the 9th district (an area of 3 km<sup>2</sup>), only 9 one-way differences were detected out of the 366 GIP one-way streets that are registered in this district. The majority



of these differences is due to errors in OSM. 5 of the 7 OSM errors were caused by the license change in July 2012 (OSM Wiki 2013) where 1 % of OSM data had to be deleted and afterwards added again by community members. The two remaining one-way differences are GIP errors. Additionally, 17 turn restriction differences out of 103 GIP turn restrictions were detected in the 9th district. 11 of these differences are due to errors in OSM, 6 are errors in GIP. These values indicate that both OSM and GIP data for the 9th district are of good quality.

In the western part of the 17th district around the train station Hernals, 20 differing one-way streets were detected on the much smaller area of 0.7 km<sup>2</sup>. Traditionally, OSM data quality shows an urban-rural divide (Thaller 2009, Zielstra & Zipf 2010) which is already noticeable in the outer districts of Vienna. 13 differences are due to errors in OSM and they have not been caused by the license change. Instead, one-way information for these streets has always been missing. The GIP export also contains 7 errors, a relatively high number as well.

Analyses showed that the Stuwerviertel is a hotspot for differing turn restrictions (21 in an area of 0.5 km<sup>2</sup>). The on-site evaluation showed that 13 differences were caused by an error in generating the OSM routing graph: Barrier tags blocking through-traffic for motorized vehicles were not considered. The remaining 8 differences were errors in the GIP export.

## 4 Conclusion and future work

This work gives an overview of the current state of street network comparisons with a focus on comparing OSM and other free and commercial street network data sets. We present common performance indicators used in related studies and propose two novel indicators for comparing turn restrictions and one-way streets.

We use indicators for positional accuracy, network length and attribute completeness in a case study comparing OSM and the official Austrian reference graph GIP. Results show that both street networks are of comparable length, and comparisons of motorway geometries attest high positional accuracy. Comparisons of the availability of street name and speed information show better levels of completeness in the GIP export.

We further assessed OSM one-way street and turn restriction data quality using the GIP export as a reference graph. Comparisons show that simply counting turn restriction objects leads to a worse quality estimate than provided by our novel routing-based turn restriction comparison. The on-site checks confirmed that our methods for one-way street and turn restriction quality assessment enable us to pinpoint areas where further investigation has to be directed in order to improve graph quality. In both graphs, most errors affect low-ranking streets. Since routing applications are expected to favour high-ranking streets, these error should only have minor effects on vehicle routing.

The analysis tools presented in this study have been developed as Sextante for QGIS geoprocessing models, using existing functionality as well as new scripts

created for this study. The tools can be used to compare street networks world-wide and are not limited to OSM comparisons. Their usage requires no software development knowledge and with all tools available on Github under open source licenses, all processing steps can be validated. Assuming familiarity with OSM and the respective reference street network, we roughly estimate that resources of two to three person weeks are necessary to reproduce this study for a region of similar size.

While this work focuses on comparing street networks for motorized traffic, the tools we present are not limited to these streets and future work will include the evaluation of graphs for other modes of transport such as biking or walking.

## 5 Acknowledgements

This work was conducted within the project EMPORA 2, funded by the Austrian Climate and Energy Fund of the Federal Ministry of Transport, Innovation and Technology within the "Electric Mobility Flagship Projects" program.

## References

- Ather, A. (2009), A Quality Analysis of OpenStreetMap Data, Master's thesis, University College London.
- Ciepluch, B., Mooney, P., Jacob, R., Zheng, J. & Winstanely, A. C. (2011), 'Assessing the Quality of Open Spatial Data for Mobile Location-based Services Research and Applications', *Archives of Photogrammetry, Cartography and Remote Sensing* **22**, 105–116.
- Dijkstra, E. W. (1959), 'A note on two problems in connexion with graphs', *Numerische Mathematik* **1**, 269–271.
- Geofabrik (2013). (accessed 2012-12-18).  
**URL:** <http://www.geofabrik.de/data/download.html>
- Goodchild, M. F. & Hunter, G. J. (1997), 'A Simple Positional Accuracy Measure for Linear Features', *International Journal of Geographical Information Science* **11(3)**, 299–306.
- Haklay, M. (2010), 'How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets', *Environment & Planning B: Planning and Design* **37**, 682–703.
- Hausdorff, F. (1914), *Grundzüge der Mengenlehre*.
- Koukoletsos, T., Haklay, M. & Ellul, C. (2012), 'Assessing Data Completeness of VGI through an Automated Matching Procedure for Linear Data', *Transactions in GIS* **16(4)**, 477–498.

- Ludwig, I., Voss, A. & Krause-Traudes, M. (2011), A Comparison of the Street Networks of Navteq and OSM in Germany, *in* 'Advancing Geoinformation Science for a Changing World', Springer, pp. 65–84.
- Neis, P., Zielstra, D. & Zipf, A. (2012), 'The Street Network Evolution of Crow-sourced Maps: OpenStreetMap in Germany 2007-2011', *Future Internet* **4**, 1–21.
- OSM Wiki (2013), 'Open Database License'. (accessed 2013-01-10).  
**URL:** [http://wiki.openstreetmap.org/wiki/Open\\_Database\\_License](http://wiki.openstreetmap.org/wiki/Open_Database_License)
- Quantum GIS Development Team (2013), 'Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project'. (accessed 2013-01-10).  
**URL:** <http://qgis.osgeo.org>
- Royal Society, T. (2012), 'Science as an open enterprise - The Royal Society Science Policy Centre report'.  
**URL:** [http://royalsociety.org/uploadedFiles/Royal\\_Society\\_Content/policy/projects/sape/2012-06-20-SAOE.pdf](http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf)
- Sextante (2013), 'SEXTANTE spatial data analysis library and geoprocessing framework'. (accessed 2013-01-10).  
**URL:** <http://sextantegis.com/>
- Thaller, D. (2009), Die Open-Source-Plattform "OpenStreetMap", eine Konkurrenz für Geodatenhersteller? Gezeigt im Vergleich mit Geodaten der Firma Tele Atlas, Master's thesis, Universität Wien.
- Zielstra, D. & Hochmair, H. H. (2011), A Comparative Study of Pedestrian Accessibility to Transit Stations using Free and Proprietary Network Data, *in* 'Proceedings of Transportation Research Board - Annual Meeting, Washington, DC, USA, 23 - 27 January 2011'.
- Zielstra, D. & Hochmair, H. H. (2012), Comparison of Shortest Path Lengths for Pedestrian Routing in Street Networks Using Free and Proprietary Data, *in* 'Proceedings of Transportation Research Board - Annual Meeting, Washington, DC, USA, 22 - 26 January 2012'.
- Zielstra, D. & Zipf, A. (2010), A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany, *in* 'AGILE 2010. The 13th AGILE International Conference on Geographic Information Science. Guimaraes, Portugal'.